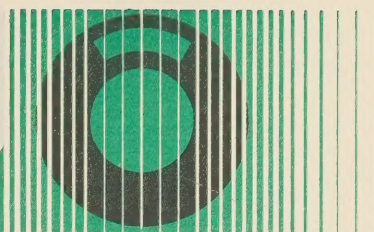




Machine Readable Archives

CAI
AK
-M19



BULLETIN

ISSN 0821-3658

Organizational Changes

The Senior Management Committee of the Public Archives of Canada announced organizational changes in the department in the fall of 1986. One of the major changes was the integration of the Machine Readable Archives Division and the Federal Archives Division into the Government Archives Division. The major objective of the reorganization is to strengthen the archival functions of acquisition, custody and reference service for government records regardless of medium. The combined resources of the two divisions will permit the appraisal and acquisition of historically valuable government records in the most appropriate form. The focus on machine readable records, their acquisition, processing and servicing remains an important priority of the Public Archives.

Acquisition of private sector data files will continue with the eventual transfer of responsibility for those records from the Government Archives Division to the Manuscript Division over the next two years.

The new organizational structure became effective on December 1, 1986. The *Machine Readable Archives Bulletin* will continue to provide information on activities relating to the creation, use and dissemination of machine readable data. The name of the publication, reflecting the new organizational structure, will be changed in the next volume.

The Appraisal of Government EDP Records

The process followed for the appraisal of EDP records is similar to that used in the appraisal of paper records in that the records are evaluated for their informational, evidential and legal value. The major difference is that not only must the content be examined, but the technical aspects associated with the records must be evaluated to ensure that they are accessible. Secondary sources are vital to the appraisal process because the records cannot be physically reviewed by the archivist in the same way as a traditional medium such as paper. All information associated with the data (inputs, outputs and documentation) is reviewed to ensure the preservation of the most useful records for research use. Certain factors are of prime importance in the evaluation of the records, such as the level of aggregation, the format, the linkage possibilities, the restrictions on access and the

frequency of change. The hardware and software managing the records can create difficulties in the future use of the data, and must be examined carefully to ensure that the most appropriate storage format is adopted.

To illustrate the appraisal process, this article will provide a detailed evaluation of the HOMEPLAN data base created in the Department of Energy, Mines and Resources.

HOMEPLAN APPRAISAL

In order to ensure that all aspects of the data are examined, the appraisal follows two major steps: content analysis and technical analysis. An overall evaluation is then provided.

Content Analysis

In 1980, the Government of Canada established the Canadian Oil Substitution Program as an integral component of its National Energy Program. Under the Canadian Oil Substitution Program (COSP), grants of up to \$800 were available to householders who substituted for oil. However, because oil substitution opportunities were limited in Prince Edward Island, Newfoundland, the Yukon and the Northwest Territories, householders, landlords and tenants could use the grant for conservation measures such as heating system retrofit, insulation, weatherstripping and caulking. Partly in conjunction with this added conservation thrust, EMR began considering an on-site energy analysis and advisory service for residential buildings, to provide guidance on how to assess the most cost-effective and optimal energy conservation measures.

In 1981-1982, EMR sponsored a program of research into energy analysis methodology and procedures. Four pilot programs were run in the areas mentioned above, and on-site audits were offered to householders. In 1983-1984, EMR continued further developmental work in Prince Edward Island, Newfoundland and the Yukon. This work saw the delivery of over 4,600 audits by regional management firms on contract to EMR throughout these three areas. It is this audit, called HOMEPLAN, that concerns us here.

During the program a considerable amount of data were generated from two sources: (1) the HOMEPLAN data form, which contains information required by the HOMEPLAN energy audit program, and (2) management reports, submitted by the regional management firms, which contain auditor-generated data. These data are held

by EMR within the Management Reporting and Technical Data Bases associated with the program. The Management Reporting Data Base was developed to collect information relating to the monthly operations of the regional management firms. The Technical Data Base is composed of the actual energy audits of houses.

The information contained in the Management Reporting Data Base was generated in the field by auditors and regional management firm personnel. In essence, the Management Reporting Data Base contains the operational/household records associated with the program. Its major variables include: expense information, completed and incomplete audit data, public demand for audit information, promotional information about the program, audit activity data, auditor performance information, householder data and audit time frame/turnaround data. All of this information is summarized in monthly reports residing in EMR's Computer Science Centre.

On the other hand, the Technical Data Base is composed of the data generated by the auditors during the actual audits. It also includes all of the householder information included in the Management Reporting Data Base. The raw data includes all of the information contained on the HOMEPLAN data form. These data are augmented by information generated during the audit by software, including information on energy savings upon retrofit. Specific data include: house type and general construction information; specific information about the house's construction, including window area data; measurements of the walls, attic, roof, foundation and floors; data on air tightness, siding, interior and exterior sheathing, insulation type and whether the house experiences moisture problems; tenure/occupancy data; primary, secondary and tertiary heating source information; and data on the house's heating system.

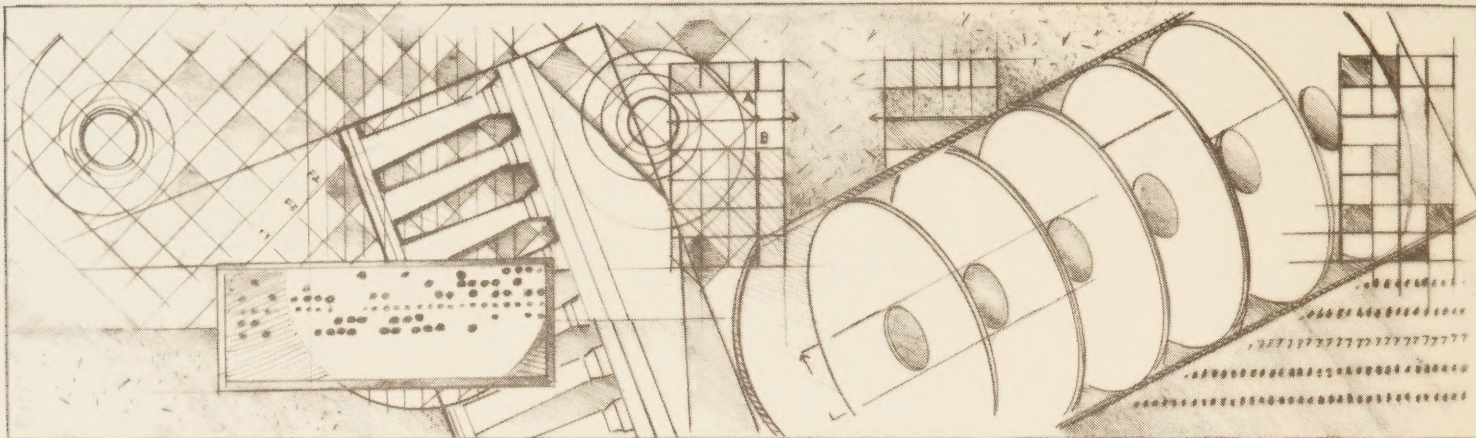
Technical Analysis

The data are stored and manipulated by a data base utility, SIR (Scientific Information Retrieval). The SIR/DBMS is maintained on Energy, Mines and Resources' CDC Cyber, which operates under NOS. The data are fully documented, and a LIST SCHEMA command function, under SIR, will provide detailed information about each variable name, the variable type (integer, floating point, string), the position on the input record (that is, data unit size), short and long

cont'd on page 2



cont'd from page 1



variable information, the range of valid values for each variable, the read and write security levels and the missing value designators.

The data base has been stored in a format similar to that of an SPSS system file and, as a result, is dependent upon SIR/DBMS for any data manipulation. In fact, SIR has been created in order that SPSS commands can be used to manipulate the data.

Evaluation

It is suggested that the division archive the Technical Data Base and not the Management Reporting Data Base, because all of the valuable information contained in the latter also resides in the former.

EMR will provide complete system documentation and SCHEMA information in

order to archive the data base. However, because some information will be lost when the Technical Data Base is down-loaded into a series of "flat files," it probably should be maintained in a SIR format. Some information will be lost because some of the variables have been created after data input, and will be lost on data output. Thus the data base should probably be maintained in CDC Cyber format and operated under an account at EMR's computer facility.

Also, the Technical Data Base should be maintained in this format for a period not to exceed five years, as EMR is going to support the SIR/DBMS only until 1991. At this time EMR will be changing from a "batch environment" to a "virtual environment," and probably will not continue to support SIR. As a result, over the next five years, the Machine Readable Archives should closely monitor the use of the data base and determine exactly

which data should be kept, and in what format. Since EMR and the MRA do not have the resources to allocate to the full-time archiving of the data base, ongoing work should be undertaken to standardize data access over the next five years. A number of tests should be undertaken to determine the exact cost of unloading those parts of the data base containing extractable information. Until this is done, routine data extraction and researcher requests should be handled through the SIR software at EMR's Computer Science Centre, on a cost recovery basis.

David Brown and
Katharine Gavrel
Machine Readable Archives Division

Anyone wishing information about the *Bulletin* may write to: Public Archives of Canada, Assistant Director, Government Archives Division, 395 Wellington Street, Ottawa, Ontario K1A 0N3, or phone (613) 996-5626.

Analyse technique

Les données sont mémorisées et mani-

puées par le programme utilisateur SIR (Scien-

tific Information Retrieval). La maintenance

du SIR/SCBD est assurée par l'ordinateur

CDC Cyber d'Énergie, Mines et Ressources

qui fonctionne à l'aide du système d'exploita-

tion NOS. Les données sont bien documen-

tées et la commande LIST SCHEMA du SIR

permet d'obtenir des renseignements détaill-

és sur chacune des variables, leur type (nom-

bre entier, décimale flottante, séquence), la

position des données (c'est-à-dire la lon-

gueur de l'unité de données), les variables

complètes ou abrégées, l'échelle de valeurs

correctes pour chaque variable, le niveau de

l'introduction et l'introduction de valeurs man-

quantés.

La base de données a été stockée sur un

format semblable à celui du fichier SPSS et la

manipulation des données doit donc être

effectuée par le SIR/SCBD. En fait, le SIR a été

conçu de manière que les commandes du

SPSS puissent servir à traiter les données.

Évaluation

La division devrait procéder à l'archi-

vage de la base de données techniques car

toutes les informations utiles de la base de

données de gestion s'y trouvent également.

Énergie, Mines et Ressources fournira

une documentation complète sur le système

et sur la commande SCHEMA pour l'archi-

vage de la base de données. Comme des

variables ont été créées une fois l'introduc-

tion des données terminée certaines informa-

tions seront éliminées lors du téléchargement de la base de données techniques dans

une série de fichiers bidimensionnels. Il

serait donc utile de la conserver sur format

SIR et de l'exploiter à l'aide du CDC Cyber

dans le cadre d'un contrat avec l'installation

informatique d'EMR.

Il faudrait également que la base de don-

nées techniques soit conservée dans ce for-

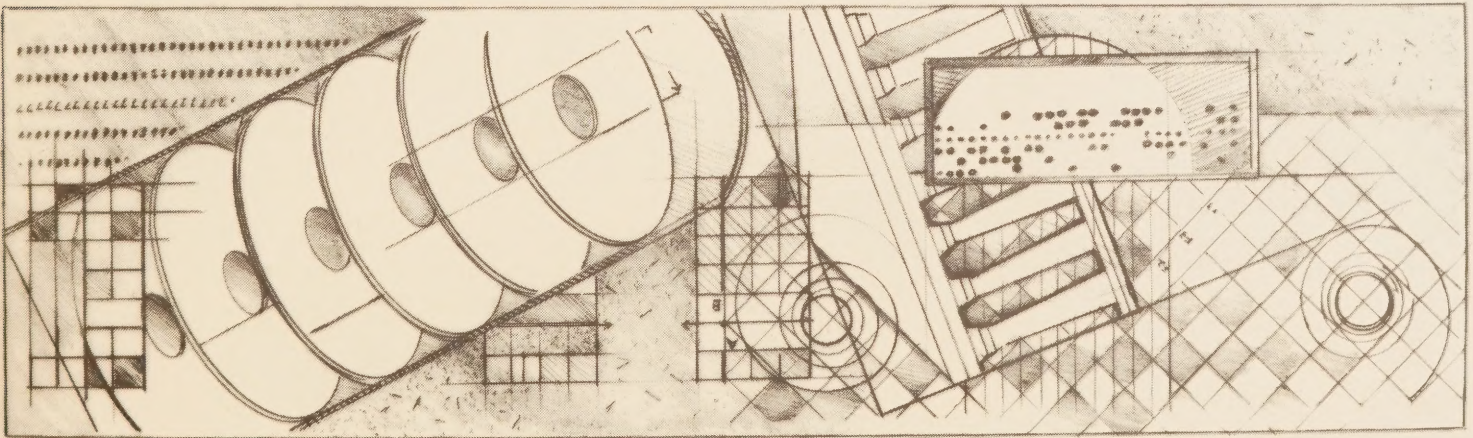
mat pour une période maximale de cinq ans

puisque le Ministère appuiera le SIR/SCBD

uniquement jusqu'en 1991 et qu'à ce moment-

là, il passera de la technique de traitement par

lots à la technique de traitement virtuel.



Au cours des cinq prochaines années les

Archives ordinales (AO) devront donc

contrôler de près l'utilisation de la base de

données et déterminer exactement quels ren-

seignements et quels formats devront être

gardés. Comme EMR et les AO n'ont pas les

ressources voulues pour travailler à plein

temps à l'archivage de la base de données, on

essayera de normaliser en permanence

l'accès aux renseignements durant cette pé-

riode. Plusieurs tests devront être réalisés

pour déterminer ce qu'il en coûtera exacte-

ment pour vidier les parties qui contiennent

des renseignements pertinents. D'ici là

l'extraction des données et le traitement des

demandes des chercheurs s'effectuera

à l'aide du programme SIR du Centre des

services d'ordinateur d'EMR, et ce, à prix

coûtant.

David Brown

Katharine Gavrel

Division des archives ordinales

Pour obtenir des renseignements au sujet du Bulletin,

veuillez écrire à : Archives publiques du Canada, Direc-

teur adjoint, Division des archives gouvernementales,

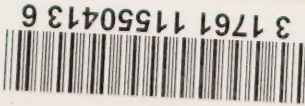
395, rue Wellington, Ottawa (Ontario), K1A 0N3, ou

téléphoner au (613) 996-5626.

©Ministre des Approvisionnement et Services Canada 1987



Archives ordinolingués



Changements

organisations

Le Comité de la haute direction des Archives publiques du Canada a annoncé à l'automne de 1986 des changements organisationnels au sein du département. L'un des principaux changements a consisté dans la fusion de la Division des archives ordinolingués avec la Division des archives fédérales, qui forment maintenant la Division des archives gouvernementales. Le principal but de cette dernière réorganisation est de renforcer les fonctions archivistiques d'acquisition, de garde et de communication de renseignements exercées à l'égard des documents gouvernementaux, quel que soit leur support. Les ressources combinées de ces deux divisions permettront d'évaluer et d'acquies les documents gouvernementaux de valeur historique de la manière la plus appropriée. L'acquisition, le traitement et la préservation des documents ordinolingués continueront d'être des activités importantes des Archives publiques.

L'acquisition de fichiers de données provenant du secteur privé se poursuivra, mais la responsabilité de ces documents sera transférée au cours des deux prochaines années à la Division des manuscrits. La nouvelle structure organisationnelle est devenue effective le 1^{er} décembre 1986. Le *Bulletin* continuera de fournir des renseignements sur les activités liées à la création, l'utilisation et la diffusion des données ordinolingués. Toutefois, la publication aura désormais sa prochaine livraison un nouveau nom qui reflètera la réorganisation effectuée.

Évaluer les documents informatisés

Les méthodes d'évaluation des documents informatisés et des documents sur papier se ressemblent en ce sens que la sélection effective toujours en fonction de la valeur informative, juridique et de témoignage. La grande différence pour les documents informatisés réside dans le fait qu'il faut examiner non seulement le contenu mais aussi le support matériel pour assurer l'accessibilité des documents à long terme. Les sources secondaires sont d'une extrême importance dans le processus d'évaluation. Toutes les informations se rapportant aux données (entrées, sorties et documentation) doivent passer en revue pour s'assurer que seuls les documents les plus utiles pour les

Analyse du contenu

En 1980, le gouvernement du Canada a créé, dans le cadre de son Programme énergétique national, le Programme canadien de remplacement du pétrole (PCRP) en vertu duquel les occupants de maison pouvaient obtenir jusqu'à 800 \$ de subvention pour remplacer leur système de chauffage au mazout. Toutefois, comme à l'Île-du-Prince-Édouard, à Terre-Neuve, au Yukon et dans les Territoires du Nord-Ouest les possibilités de conversion étaient limitées, les occupants, propriétaires et locataires, pouvaient utiliser la subvention à d'autres projets de conservation de l'énergie tels que la modernisation des systèmes de chauffage, la pose d'isolant et de bandes d'étanchéité et le calfeutrage. En partie à la suite de cette initiative, EMR a commencé à envisager la création de services d'analyse et de consultation qui proposeraient les moyens les plus efficaces pour économiser l'énergie dans les logements.

En 1981-1982, le Ministère a parrainé un programme de recherche sur les méthodes et les autres de l'analyse énergétique. Dans les provinces et territoires susmentionnés, quatre projets pilotes ont été réalisés, et les occupants se sont vus offrir des services de vérification sur place. L'année suivante, EMR a poursuivi ses travaux de développement à l'Île-du-Prince-Édouard, à Terre-Neuve et au Yukon. Plus de 4 600 vérifications ont été réalisées par des sociétés de gestion régionales dont le Ministère avait retenu les services.

C'est de ces vérifications dans la base de données HOMEPLAN dont il est question. Enormément de données ont été produites dans le cadre du programme grâce à deux outils : les formules de saisie HOMEPLAN qui comprennent les renseignements demandés lors des vérifications des bilans énergétiques, et les rapports de gestion présentés par les sociétés régionales qui contiennent les données consignées par les vérificateurs. Le Ministère conservait ces renseignements soit dans une base de données de gestion. Cette dernière avait été créée pour documenter les activités mensuelles des sociétés de gestion régionales, tandis que la base de données techniques comprenait les vérifications du rendement énergétique proprement dites. Les informations stockées dans la base de données de gestion étaient générées sur place par les vérificateurs et les employés des sociétés régionales. La base de données contenait essentiellement des documents sur l'administration et la gestion interne. On y trouve des renseignements sur les dépenses, les vérifications complètes ou incomplètes, les demandes de renseignements présentées par le public, la publicité, les activités de vérification, le rendement du vérificateur, l'occupant et le calendrier ou le délai d'exécution. Toutes ces informations sont réunies dans des rapports mensuels conservés au Centre des services d'ordinateur d'EMR.

La base de données techniques comprend des données ayant trait à l'économie d'énergie réalisée après la modernisation du système de chauffage. La base de données renferme aussi des renseignements généraux sur le type de maison et la construction, des données précises sur le domicile, entre autres des autres des fenêtres et les dimensions des murs, du grenier, du toit, des fondations et des planchers, des informations sur l'étanchéité, le bardage, le revêtement intérieur et extérieur, les isolants et les problèmes d'humidité, des données sur le bail et l'occupation de la maison et sur les sources de chauffage primaire, secondaire et tertiaire, ainsi que des renseignements techniques sur le système de chauffage.